

The Distribution of Transgene Insertion Sites in Barley Determined by Physical and Genetic Mapping

Haroldo Salvo-Garrido,^{1,2} Silvia Travella,^{1,3} Lorelei J. Bilham,¹ Wendy A. Harwood⁴
and John W. Snape

John Innes Centre, Department of Crop Genetics, Norwich Research Park, Norwich NR4 7UH, United Kingdom

Manuscript received November 6, 2003

Accepted for publication April 2, 2004

ABSTRACT

The exact site of transgene insertion into a plant host genome is one feature of the genetic transformation process that cannot, at present, be controlled and is often poorly understood. The site of transgene insertion may have implications for transgene stability and for potential unintended effects of the transgene on plant metabolism. To increase our understanding of transgene insertion sites in barley, a detailed analysis of transgene integration in independently derived transgenic barley lines was carried out. Fluorescence *in situ* hybridization (FISH) was used to physically map 23 transgene integration sites from 19 independent barley lines. Genetic mapping further confirmed the location of the transgenes in 11 of these lines. Transgene integration sites were present only on five of the seven barley chromosomes. The pattern of transgene integration appeared to be nonrandom and there was evidence of clustering of independent transgene insertion events within the barley genome. In addition, barley genomic regions flanking the transgene insertion site were isolated for seven independent lines. The data from the transgene flanking regions indicated that transgene insertions were preferentially located in gene-rich areas of the genome. These results are discussed in relation to the structure of the barley genome.

ONE of the main goals of plant genetic engineering is to achieve stable transgenic events that give predictable and reproducible levels of expression of a transgene and that are fully characterized in terms of the effect and implications of the transgene insertion for the plant. To attain this goal, it is necessary to obtain detailed information about the transgene insertion site and to characterize transgenic plants in detail at the physical, genetic, and molecular levels. An understanding of the transgene insertion site within the host genome is an essential first step toward understanding the importance of the transgene insertion position and in aiding the safety assessment process for a particular transgenic event. The physical position of transgenes has been detected by fluorescent *in situ* hybridization (FISH) in cereals such as barley, wheat, triticale, and oat (PEDERSEN *et al.* 1997; LEGGETT *et al.* 2000; SVITASHEV *et al.* 2000; SALVO-GARRIDO *et al.* 2001; CHOI *et al.* 2002). However, equally important is the genetic map position of transgenes, and to our knowledge, this has not been studied in any crop species. Combining knowledge of

the physical position of transgenes with their genetic map position provides confirmation of the insertion site and a better understanding of the pattern of transgene insertions. It is also a valuable starting point for understanding the genomic environment in which a transgene resides and for determining whether transgene insertion is indeed a random process.

Additional information on transgene insertions can be obtained from the analysis of transgene flanking regions. Large-scale studies of T-DNA insertions have been undertaken in *Arabidopsis* using this method (FORSBACH *et al.* 2003; QIN *et al.* 2003a,b) and also in rice (SHA *et al.* 2004). Such studies are leading to an increased understanding of the preferences for transgene insertion and its implications. However, this has not yet been done for the large-grained cereals, such as barley, where ~80% of the genome is repetitive DNA.

In the present study, the physical position of transgenes in 19 independent transgenic barley lines has been determined using a highly sensitive and reproducible method of determining the chromosomal position of transgenes on metaphase chromosomes (SALVO-GARRIDO *et al.* 2001). These locations were then confirmed by genetic mapping of the transgenes in 11 of these independently transformed lines. This is the first report of both physical and genetic mapping of transgenes in cereals. In addition, further information on the transgene insertion environment was obtained by isolating and characterizing the genomic sequence flanking the transgenes in some of the lines. The patterns of trans-

¹These authors contributed equally to this work.

²Present address: Biotechnology Unit, INIA, Carillanca, P.O. Box 58 D, Temuco, Chile.

³Present address: Institute of Plant Biology, University of Zurich, 8008 Zurich, Switzerland.

⁴Corresponding author: John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, United Kingdom.
E-mail: wendy.harwood@bbsrc.ac.uk

TABLE 1
The transgenic lines used for physical and genetic mapping

Parents		Transgenic generation	Constructs	Gene transfer methods	Chromosomal location of insertion
Cross parent	Transgenic line				
Nigrinudum	AGN1A	T ₁	pDM805	AMT	2H
	AGN4A	T ₀	pDM805	AMT	5H
	HB1A	T ₁	pAHC25	PB	5H
	HC1B	T ₂	pAL51 + pAL74	PB	4H
Hayakeizo	HC2B	T ₃	pAL51 + pAL74	PB	5H
	HC8	T ₁	pAL51 + pAL74	PB	4H
Nigrinudum	HC12B	T ₂	pAL51 + pAL74	PB	6H
Nigrinudum	HG4B	T ₀	pAL51 + R173	PB	2H
Nigrinudum	HG3B	T ₀	pAL51 + R173	PB	4H
Nigrinudum	HI6A	T ₁	pAL51 + pDAPH7	PB	6H
	HI8A	T ₂	pAL51 + pDAPH7	PB	6H
	HI9A	T ₁	pAL51 + pDAPH7	PB	5H
	HI1A	T ₁	pAL51 + pDAPH7	PB	3H
	HH2A	T ₁	pAL51 + pPhyt	PB	5H
	HH3E	T ₁	pAL51 + pPhyt	PB	6H
	HH4F	T ₂	pAL51 + pPhyt	PB	5H
	HR18D	T ₀	pAL51 + γ -gliadin	PB	4H
	Tr1372D8	T ₁	pAHC25 + pDC-2CP	PB	4H
	Tr130	T ₁	pAHC25 + pDC-2CP	PB	2H

Details of crosses made, constructs, transgenic generation, gene-transfer method, and chromosomal location for each transgenic line are included. PB, particle bombardment; AMT, agrobacterium-mediated transformation.

gene insertion observed are discussed in relation to patterns of insertion seen in other plants and to the structure of the barley genome.

MATERIALS AND METHODS

Plant materials: The plant materials studied by FISH and mapping procedures were T₁ transgenic plants and F₂ populations of crosses between T₀, T₁, T₂, or T₃ transgenic lines of barley (*Hordeum vulgare* cv. Golden Promise) with the non-transgenic barley varieties, Nigrinudum, Hayakeizo, Fanfare, and Bronze. Table 1 summarizes the crosses, constructs, and gene transfer methods used. All transgenic lines were produced at the John Innes Centre except lines Tr130 and Tr1372D8, obtained from Monsanto (Cambridge, United Kingdom). Most of the lines were produced by particle bombardment of immature embryos (HARWOOD *et al.* 2002), except for lines AGN1A and AGN4A, which were transformed by Agrobacterium following the method of TINGAY *et al.* (1997). The plasmid pPhyt contained a phytase gene under the control of an α -amylase promoter, pDAPH7 contained the *DapA* gene involved in lysine biosynthesis, pAL74 contained a fungal glucoamylase gene under an α -amylase promoter, and pDC-2CP contained a viral coat protein. All other constructs—pAHC25 (CHRISTENSEN and QUAIL 1996), pAL51 (HARWOOD *et al.* 2002), and the binary vector pDM805 (TINGAY *et al.* 1997)—contained the *bar* gene as a selectable marker, which encodes resistance to the glufosinate group of herbicides, together with a reporter gene, either the *gus* gene (β -glucuronidase) or the firefly luciferase gene. All transgenic lines examined expressed the *bar* gene.

Physical mapping by FISH: FISH was performed according to SALVO-GARRIDO *et al.* (2001). Simultaneous hybridization

was carried out with probes pTa71 (18S-5.8S-26S rDNA) and pTa794 (5S rDNA), which hybridize to ribosomal DNA to give distinct and specific chromosomal patterns that unambiguously identify all seven individual chromosomes in barley (LEITCH and HESLOP-HARRISON 1993).

Marker probe pTa71 hybridized to specific sites of ribosomal DNA on the short arms of chromosomes 1H, 2H, 5H, 6H, and 7H. Marker probe pTa71 labels chromosomes 5H and 6H at the nucleolar organizer region (NOR) on the short arm to produce the strongest hybridization bands across both chromatids. Chromosomes 1H, 2H, and 7H are more weakly labeled by pTa71, but show a specific hybridization signal in each chromatid. It is possible to identify chromosomes 5H and 6H without a marker probe on the basis of morphology alone because of the presence of the NOR, which is visible as a secondary constriction in a chromosome spread. These chromosomes can be distinguished by observing that chromosome 5H has a much longer, long arm compared to its short arm whereas chromosome 6H has a short and long arm of equal length.

Marker probe pTa794 hybridized to regions of ribosomal DNA in the long arm of chromosomes 2H, 3H, and 4H and to a region of ribosomal DNA in the short arm of chromosome 7H. The strongest signal following hybridization of pTa794 was found on chromosome 2H. To identify chromosome 4H, the hybridization signal was always found at the very end of the long arm. Chromosomes 3H and 7H could be distinguished due to differences in chromosome morphology and the hybridization of pTa794 to regions in the long arm of chromosome 3H and the short arm of chromosome 7H.

Mapping procedures: Bulk segregant analysis (BSA) was initially applied to establish marker/transgene associations (MICHELMORE *et al.* 1991). F₂ plants at the three/four leaf stage, derived from the crosses described above, were pheno-

typed in terms of resistance and susceptibility to the herbicide. For BSA analysis, DNA of the resistant plants was bulked into one sample. DNA of susceptible plants was isolated individually and a bulked sample was also prepared. *Bar* PCR was performed on the sensitive plants to check for transgene silencing and any PCR-positive plants were discarded from the mapping population.

Molecular markers: Simple sequence repeat (SSR) and restriction fragment length polymorphism (RFLP) assays were employed. Three different sources of microsatellite primers were used (BECKER and HEUN 1995; LIU *et al.* 1996; RAMSAY *et al.* 2000). The RFLP probes were those already widely mapped in barley and wheat.

Linkage analysis and estimation of recombination frequency: The physical mapping results provided a starting point for genetic mapping by locating transgenes to chromosomes and chromosome arms, as described above. Known polymorphic codominant markers for the target region were then genotyped on the susceptible population. Two-point recombination frequencies were obtained from a simple maximum-likelihood estimation equation and map order was calculated. Recombination frequency was converted to centimorgans using the Kosambi mapping function. Comparative mapping with published maps of barley for the corresponding regions was used to compare gene orders.

Isolation of transgene flanking regions: The lines selected for analysis of transgene flanking regions were all transformed with plasmid pAL51, containing the luciferase (*luc*) reporter gene. All lines had a low copy number and simple integration patterns as determined by Southern analysis. In addition, they did not contain an intact copy of the luciferase gene as the plasmid breakpoint was within the luciferase cassette. Genomic DNA was extracted using a CTAB protocol (HARWOOD *et al.* 2000). PCR was carried out using a range of primer combinations within the *luc* cassette. In this way the breakpoint within the *luc* cassette was identified to within 200–550 bp for each line. Adaptor-mediated PCR was then used to allow the specific amplification of the transgene flanking region. The method described by SPERTINI *et al.* (1999) was used with the following modifications. PCRs were performed in a 25- μ l total reaction volume using 2 μ l ligated DNA, 1 μ l 2.5 mM nested primer, 1 μ l 2.5 mM AP1/AP2 (SPERTINI *et al.* 1999), and 21 μ l ABgene 1.1 \times Reddy Mix with 1.5 mM MgCl₂. The second PCR amplification was performed using 2 μ l of product from the first PCR amplification. Following the second PCR amplification, the total PCR product was run out on a 1% agarose gel and bands were excised. DNA was extracted from the gel bands using a QIAGEN (Valencia, CA) Qiaquick gel extraction kit. The extracted DNA was cloned using the Promega (Madison, WI) pGem-T Easy Vector System I, prior to Perkin-Elmer (Norwalk, CT) Big Dye terminator version II sequencing. Sequence data were analyzed by BLAST against the National Center for Biotechnology Information (NCBI; <http://www.ncbi.nlm.nih.gov>) and Dupont databases.

Statistical analyses: An exact chi-square test (using StatExact 4) was carried out to determine whether the transgene insertions were uniform over the barley genome. The different sizes of the barley chromosomes were taken into account during the analysis.

RESULTS

Physical map position of transgenes: The improved protocol described by SALVO-GARRIDO *et al.* (2001) was used for the detection of transgenes by FISH. This method produced specific signals on metaphase chro-

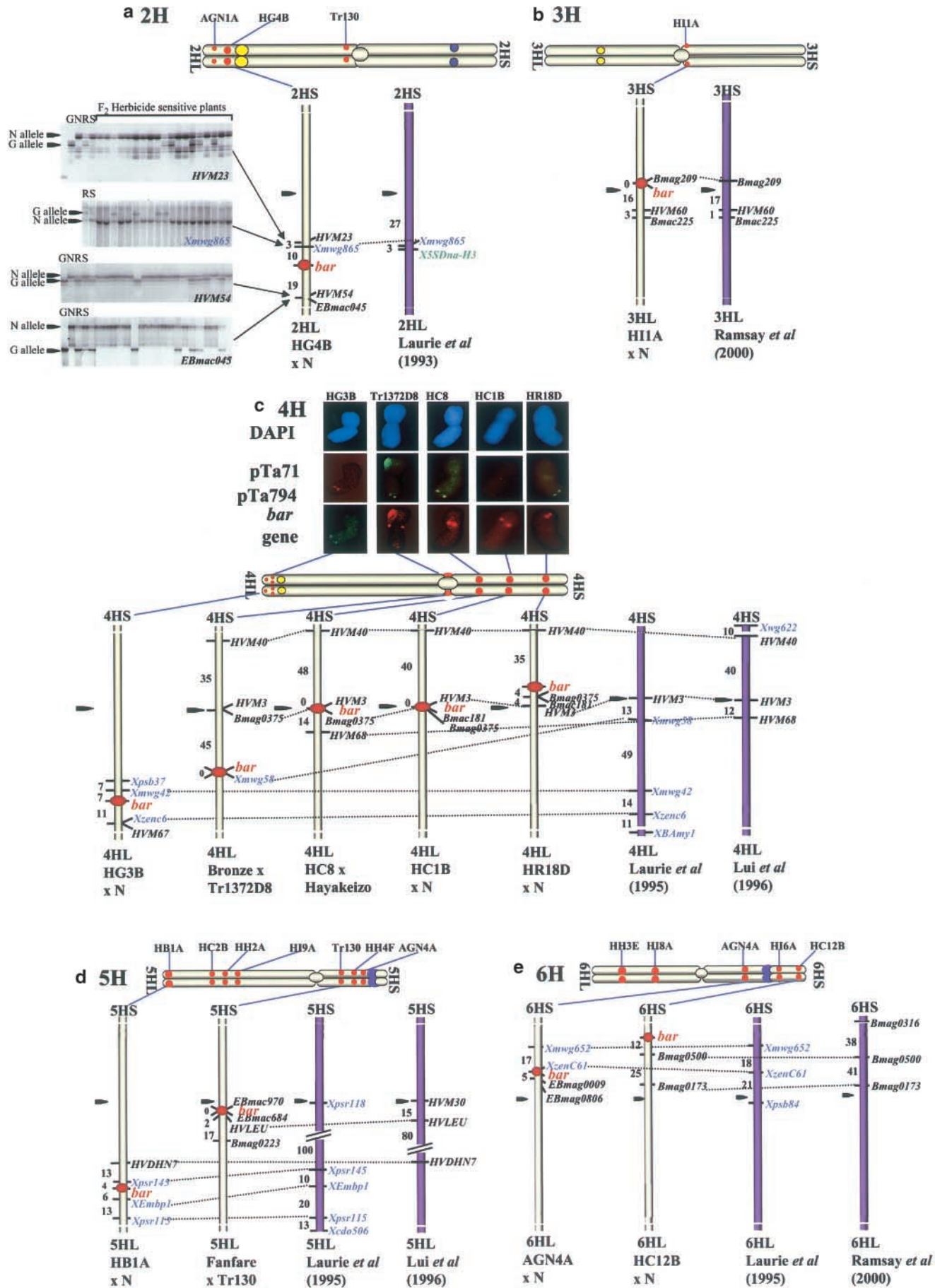
mosomes in all the lines under study. Simultaneous hybridization with karyotyping ribosomal probes (pTa71 and pTa794) allowed the clear identification of the seven barley chromosomes.

The physical location of transgenes was determined in 19 independent barley lines, and a total of 23 transgene integration sites were detected (Figures 1 and 2). Lines Tr130 and AGN4A showed two insertion sites on separate chromosomes; line HG3B showed three insertions on the same chromosome (SALVO-GARRIDO *et al.* 2001) while the remaining lines showed a single insertion site. Of the seven barley chromosomes, five (2H, 3H, 4H, 5H, and 6H) were involved in transgene integration, but no sites were detected on chromosomes 1H and 7H (Figure 2).

FISH detected 39% of the transgene signals as being in the telomeric or subtelomeric regions of both the long and the short arms. The centromeric regions contained 17% of the transgene signals. Other regions of the long arms of chromosomes contained 22% of the transgene signals while the short arms contained the remaining 22%. Most integration sites were found on chromosomes 5H (30%), 4H (30%), and 6H (22%). An exact chi-square test (using StatExact 4) was carried out to determine whether the transgene insertions were uniform over the barley genome (Table 2). The exact test takes into account the small sample size by conditioning the distribution of counts on the total; hence the possible arrangements of counts in the cells of the 1×7 chi-square table follow a multinomial distribution. The different sizes of the seven barley chromosomes were also taken into account in determining the expected transgene insertions for each chromosome (LAURIE *et al.* 1992). Strong evidence was found against the null hypothesis that there was no difference in the frequency of transgene insertions between the different chromosomes ($P > 0.05$). Thus it was concluded, on the basis of the FISH results, that transgene insertion was not random in the barley lines analyzed in this study.

Transgenic line Tr130 was cotransformed with constructs pAHC25 (*bar* gene) and pDC-2CP (viral coat protein). Two different integration sites were observed for pDC-2CP and only one integration site was mapped with pAHC25. On chromosome 5H, both pDC-2CP and pAHC25 shared the same integration site close to the centromere in chromosome 5H, while pDC-2CP gave an additional signal centromeric in chromosome 2H. These results indicate that the *bar* and the viral coat protein gene have integrated independently into the genome.

Genetic map position of the transgenes: The genetic location of transgenes was determined for 11 independent lines (Figure 1). All genetic maps showed collinearity to published SSR and RFLP barley maps. The physical integration sites showed complete agreement with the genetic map positions of the transgenes. The length distortion between the genetic and physical distance of



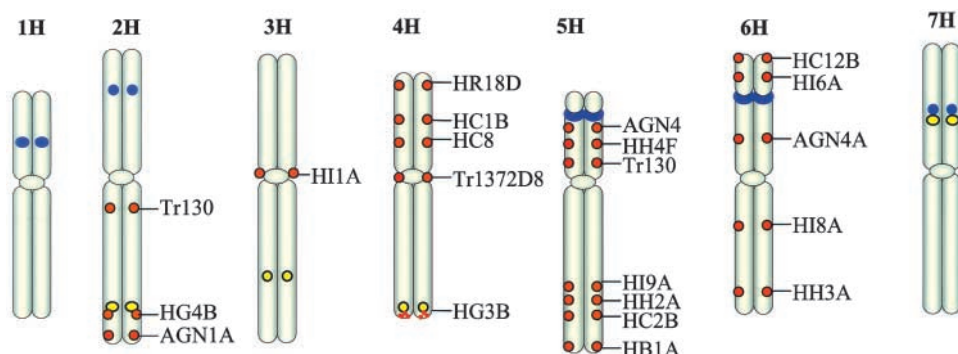


FIGURE 2.—Schematic of the physical distribution of transgene insertion sites on the seven barley chromosomes. Integrated transgenes (red spots) and the physical locations of 5S rDNA, pTa794 (in yellow), and the 18S-5.8S-26S rDNA, pTa71 (in blue) are indicated.

the transgenes is similar to that previously observed within cereal species (GILL *et al.* 1996).

The transgene in line HG4B was physically mapped to the telomeric region of the long arm of chromosome 2H (Figure 1a). In the genetic map, the *bar* gene showed linkage with the RFLP marker *Xmwig895*, mapped in the long arm of chromosome 2H (LAURIE *et al.* 1993), 27 cM from the centromere and 3 cM from the 5S rDNA site (designated *5SDna-H3*; LEITCH and HESLOP-HARRISON 1993). In line HG4B, the locus *Xmwig895* was mapped 10 cM from the *bar* gene, indicating that the transgene was genetically located close to the 5S rDNA locus. This confirmed the results of the physical mapping. In the cross HI1A × Nigrinudum, SSR locus *Bmag-209* mapped in the proximal region of the short arm of 3H (RAMSAY *et al.* 2000) and showed complete linkage to the transgene (Figure 1b).

All transgene insertions found by FISH on chromosome 4H were genetically mapped (Figure 1c). FISH indicated that the transgene in cross HG3B × Nigrinudum was telomeric in the long arm of chromosome 4H, and genetic mapping showed that it was 11 cM from the markers *XzenC6* and *HVM67*. In the barley map developed by LIU *et al.* (1996), *HVM67* was mapped 8 cM from the most terminal marker *BAmy1* (KLEINHOFs *et al.* 1993), confirming this result. In the cross Bronze × Tr1372D8, the transgene was expected to be in the centromeric region of chromosome 4H on the basis of FISH results. Total linkage was found with RFLP marker *Xmwig58*, previously mapped close to the 4H centro-

meric region (KLEINHOFs *et al.* 1993). The genetic position of the *bar* gene in line Tr1372D8 was therefore in the proximal region of the long arm of chromosome 4H.

SSR markers *HVM3* and *Bmag0375* were totally linked to the *bar* gene in cross HC8 × Hayakeizo, and both have been mapped to the 4H centromeric region (LIU *et al.* 1996; RAMSAY *et al.* 2000). The transgene in the cross HC1B × Nigrinudum showed total linkage to markers *HVM3*, *Bmac181*, and *Bmag0375*, centromeric on chromosome 4H and similar to the transgene in line HC8. With respect to the cross HR18D × Nigrinudum, the two SSR markers *Bmag0375* and *Bmac181*, which cosegregated together, were mapped 4 cM from the *bar* transgene, again in the proximal region of the short arm of chromosome 4H, close to the centromere. These three independent transgene insertions therefore all mapped to within 4 cM on the short arm of chromosome 4H. The total map distance for chromosome 4H is ~120 cM, so to find three insertions within 4 cM suggests some preference for insertion into this region of the genome.

Transgene insertions in lines HB1A and Tr130 were genetically mapped to chromosome 5H (Figure 1d). The transgene in the cross HB1A × Nigrinudum was closely flanked by loci *Xpsr145* and *XEmbp1* and mapped to the terminal region of chromosome 5H (LAURIE *et al.* 1995), confirming the FISH results. SSR loci *Ebmacc-970* and *Ebmacc684* cosegregated with the *bar* gene in the cross Fanfare × Tr130, deduced to be in the centromeric region of chromosome 5H.

FIGURE 1.—Summary of the 19 physically and 11 genetically mapped transgene insertion sites found on chromosomes (a) 2H, (b) 3H, (c) 4H, (d) 5H, and (e) 6H of barley. FISH was carried out using a biotin- or digoxigenin-labeled probe, made from fragments of the integrated plasmid, together with labeled marker probe pTa71 or pTa794, to metaphase chromosomes of barley. FISH data are represented on ideograms of each chromosome (a–e) with actual results shown only for chromosome 4H (c). Integration sites are represented by red spots and the physical locations of 5S rDNA, pTa794 (in yellow), and the 18S-5.8S-26S rDNA, pTa71 (in blue) are indicated. Transgenic line names are given above the ideogram of each chromosome and linked to the site of transgene insertion. Genetic mapping results are shown below the ideogram of each chromosome and linked to the site of transgene insertion found by FISH. Published genetic maps are shown in purple with the source reference below. Genetically mapped transgene insertion sites are shown on yellow maps with the cross-reference below. The locations of SSR markers are shown in black type, RFLP markers are shown in blue type, and the site of probe pTa794 (X5S DNA-H3) is shown in green type. (a) includes SSR and RFLP genetic mapping results from the cross HG4b × N. Golden Promise is represented by G, Nigrinudum is represented by N, the bulk of resistant individuals is represented by R, and the bulk of the sensitive individuals is represented by S.

TABLE 2
Exact chi-square analysis of the distribution of transgene insertion sites

	Chromosome no. and size (10 ⁶ bp)							Totals (5352)
	1H (644)	2H (811)	3H (797)	4H (779)	5H (756)	6H (740)	7H (825)	
Observed transgene integration sites	0	3	1	7	7	5	0	23
Expected transgene integration sites	2.77	3.49	3.43	3.35	3.25	3.18	3.55	23.000
Chi square	2.77	0.07	1.72	3.98	4.33	1.04	3.55	17.455

Transgene insertions from two independently transformed lines were mapped to the short arm of chromosome 6H (Figure 1e). The transgene in cross AGN4A × *Nigrinudum* was physically mapped to the middle of the short arm. RFLP marker *XzenC61* was totally linked to the transgene, confirming the FISH result, since this marker has been mapped 21 cM from the centromeric region on the short arm of chromosome 6H (LAURIE *et al.* 1995). The transgene in the cross HC12B × *Nigrinudum* was mapped 12 cM from the SSR locus *Bmag0500* on the short arm of chromosome 6H (RAMSAY *et al.* 2000).

Transgene flanking regions: Genomic DNA flanking one side of the transgene insertion was isolated for 5 (HC1B, HC8, HI8A, HH2A, and HH4F) of the 19 mapping lines together with two additional lines (HC3 and HC4C). Comparison of the sequence data with available databases showed that the isolated genomic DNA corresponded to recognized coding regions in 6 of these lines, suggesting that the transgene insertions have occurred in coding regions of the barley genome (Table 3).

A total of 500 bp of genomic junction sequence DNA was obtained from line HC1B. The strongest alignment was found when two regions of HC1B flanking DNA matched to a cytochrome B gene; a 107-bp region showed 95% identity and a 207-bp region showed 94% identity on the basis of the results of a BLASTn search of the NCBI nonredundant database.

With respect to line HC8, analysis of 700 bp of genomic DNA flanking the transgene insertion showed that a 183-bp region matched with 90% identity to a lipase-like protein from rice on the basis of the results of a BLASTn search of the Dupont wheat expressed sequence tag (EST) database. From line HI8A, a total of 400 bp of flanking sequence was obtained. The most significant match found following BLASTn analysis of the Dupont wheat EST database was to a catalase gene. In this case a 133-bp region from HI8A flanking DNA showed 98% identity.

For line HH4F, 500 bp of junction sequence DNA was isolated. The most significant alignment was to a barley bacterial artificial chromosome (BAC) on chromosome 5H, thus further confirming the chromosomal

location of this transgene. Homology was also seen to a high-molecular-weight glutenin. In addition to this, a 267-bp region of HH4F flanking sequence showed 52% identity to a rice polypeptide following a BLASTx search of the Dupont wheat EST database. Line HH2A was the only line where homology to known coding regions was not found in the transgene flanking DNA. Good homology was found to another barley BAC and short regions of the 500-bp flanking region showed homology to a putative retroelement from *Arabidopsis*. The two additional lines, HC3 and HC4C, showed homology to a probable glucosidase protein and a tomato leaf-curl-disease-associated sequence, respectively.

DISCUSSION

Although there have been previous reports of the physical mapping of transgenes using FISH, this is, to our knowledge, the first report of the combined physical and genetic mapping of transgenes in a crop plant. For the 19 lines analyzed in this study, only five of the seven barley chromosomes (2H, 3H, 4H, 5H, and 6H) were found to contain transgene integration sites. In addition, there appeared to be specific regions of chromosomes 4H and 5H where clusters of transgene insertions were seen. The distribution of transgene integration sites was found to statistically deviate from a random expectation. The close proximity of transgene integration sites from three independent lines on the short arm of chromosome 4H (lines HC8, HC1B, and HR18D) and also clusters of three independent transgene insertions on both the long arm (lines HC2B, HH2A, and HI9A) and the short arm (lines Tr130, HH4F, and AGN4A) of 5H suggests that these areas of the genome may be more amenable for transgene insertion than other regions of the genome. On chromosome 4H, three transgene insertions were found within a genetic distance of 4 cM of a total of 120 cM for chromosome 4H. It has been estimated that only 12% of the barley genome is occupied by the "gene space" (BARAKAT *et al.* 1997). This region is characterized by a specific GC percentage range and most genes are found within this genomic space. In wheat, genes have been found to

TABLE 3

Most significant alignments for the transgene flanking regions based on BLASTn and BLASTx searches of the NCBI nonredundant databases and the Dupont wheat and oat EST databases

Line	BLASTn results			BLASTx results		
	Alignment	Accession	E value	Alignment	Accession	E value
HC1B	Cytochrome B	AF205088	$e-111$	Cytochrome B	AAK54411	$1e-19$
HC3	No good matches			Arabidopsis putative protein (D ²)	NP_195322	$7e-05$
				Probable glucosidase protein	NP_519202	$3e-18$
HC4C	Tomato leaf-curl-disease-associated sequence	AJ316036	$3e-57$	Glycine-rich protein	CAC87058	0.12
HC8C	Rice lipase-like protein (D ¹)	AP003710	$2e-60$	Arabidopsis lipase-like protein (D ²)	NP_193590	$2e-35$
				Rice lipase-like protein	NP_917262	$3e-34$
HI8A	Catalase (D ¹) mRNA clone	X85130	$4e-62$	Catalase (D ²)	I40767	$4e-17$
		AJ309392	$8e-06$	Rice lipase-like protein	NP_917282	$5e-14$
HH2M	Barley BAC 184G9	AY268139	$6e-45$	Arabidopsis Athila retroelement (D ²)	NP_178678	$9e-06$
HH4F	Barley chromosome 5H BAC 635P2	AY013246	$2e-60$	Rice polyprotein (D ²)	AF259976	$4e-31$
	<i>Triticum turgidum</i> HMW-glutenin	AY368673	$2e-11$	Rice gag-pol precursor	AAL75983	$8e-17$

D¹, based on the results of a BLASTn search on the Dupont database. D², based on the results of a BLASTx search of the Dupont database.

occur in clusters along chromosomes (GILL *et al.* 1993), providing further evidence for gene-rich regions interspersed with large gene empty compartments in members of the Gramineae. In the present study, only lines expressing the *bar* gene were evaluated, not lines where the *bar* gene was present but not expressed. The non-random distribution of functional transgenes may therefore reflect the fact that they are found only in the 12% gene-rich region of the barley genome. The analysis of the transgene-flanking region in 7 lines supports this suggestion as most flanking regions were recognized as known coding regions when the sequences were compared to those held in databases.

Studies of T-DNA insertions in Arabidopsis (FORSBACH *et al.* 2003; QIN *et al.* 2003a,b), identified by the isolation of T-DNA flanking sequences, suggest a random distribution of the T-DNA insertions among the five chromosomes. This is perhaps not surprising, considering the differences between the genomes of cereals like barley and Arabidopsis. As previously stated, barley is thought to have large noncoding regions interspersed with gene-rich regions. The much smaller Arabidopsis genome consists of 85% gene-rich regions (BARAKAT *et al.* 1998) compared to 12% in barley, so it would not be expected that any bias would be seen. FORSBACH *et al.* (2003) examined the distribution of T-DNA copies within coding and noncoding regions of the Arabidopsis genome in further detail. They found that although the majority of the insertions were in noncoding regions, a

large number of these were found in regions immediately upstream or downstream of coding sequences. In particular, they found that T-DNA insertions in 5' upstream regions were more common than might be expected from a completely random distribution. These results, therefore, support the conclusions of the present study showing that transgene insertion may not be random.

QIN *et al.* (2003b) analyzed 1194 T-DNA insertion sites in the Arabidopsis genome by matching T-DNA flanking sequence to the Arabidopsis genome sequence. The authors observed "hot regions" that had more insertions than other areas of the genome. This result corresponds to our finding of apparent "hot spots" for transgene insertion within the barley genome. In addition, of the 1194 T-DNA insertion sites mapped, 1010 were inserted in, or close to, genes. This is further strong evidence for a preferential insertion in gene-rich regions.

BARAKAT *et al.* (2000) compared the distribution of T-DNA insertions in Arabidopsis and rice by localizing T-DNA in fractions of plant DNA separated according to their GC levels. They found T-DNA in fractions representing 85% of the Arabidopsis genome, whereas it was found only in fractions representing 25% of the genome in rice. The fractions containing the T-DNA matched those corresponding to the gene-rich region in each species. This again supports the finding of the present study that transgenes tend to insert in gene-rich regions.

SALLAUD *et al.* (2003) isolated the region flanking the left border of the T-DNA in 477 transgenic rice plants and comment that the detailed analysis of large populations of T-DNA insertion lines now in progress will provide further information on the preference of T-DNA integration in rice. A recent study by SHA *et al.* (2004) examined 361 T-DNA flanking regions in rice. These authors also detected preferential insertion of T-DNAs into protein-coding regions and an apparent nonrandom distribution of the T-DNA insertions.

CHOI *et al.* (2002) examined transgene location in 19 barley lines using FISH only. In that study, transgene insertion appeared to be random but transgene insertions were not confirmed by genetic mapping or flanking sequence data. It is therefore possible that these authors did not observe the preferred regions for transgene insertion that we detected following more detailed analysis of insertion sites.

Reports of transgene detection in barley (PEDERSEN *et al.* 1997; CHOI *et al.* 2002) and other cereal species (SVITASHEV *et al.* 2000) indicate a tendency toward the localization of transgenes in distal (subtelomeric and telomeric) chromosome regions. This was also seen in the present study where 39% of transgene insertions were found in telomeric or subtelomeric regions. This is similar to CHOI *et al.* (2002) who found 21% of insertions in telomeric regions and 10% in subtelomeric regions in barley. In oat, SVITASHEV *et al.* (2000) found 18 of 25 integration events in telomeric and subtelomeric regions. These authors suggest that this may be due to the fact that these regions are enriched with coding DNA sequences. This tendency for telomeric and subtelomeric regions to contain more transgene insertions again supports the findings of the present study. If transgene insertions are more likely in coding regions of the genome, then it is likely that some preference for telomeric and subtelomeric regions will be observed in the cereals.

In summary, information on both the physical and the genetic location of transgenes, together with data on their flanking sequences, is an important starting point toward increasing our understanding of transgene integration. Such information may also be of value in examining the effect of insertion site on transgene stability and in looking for possible unintended effects of the transgene insertion. The apparent nonrandom distribution of transgene insertions observed for the transgenic barley lines examined in this study appears to reflect the preference for insertion into gene-rich regions. Data from transgene-flanking sequences provided additional evidence for preferred insertion into gene-rich sites. This preference has also been observed in other plant species although the physical pattern of transgene insertion observed is dependent on the structure of the genome being studied.

We gratefully acknowledge the statistical assistance of Sandro Leidi of the Reading statistical service. This work was supported by the Food

Standards Agency of the United Kingdom. H. Salvo-Garrido acknowledges the support of the National Institute of Agriculture Research (INIA), Temuco, Chile, and S. Travella acknowledges the support from the European Commission (Agriculture and Fisheries program).

LITERATURE CITED

- BARAKAT, A., N. CARELS and G. BERNARDI, 1997 The distribution of genes in the genomes of Gramineae. *Proc. Natl. Acad. Sci. USA* **94**: 6857–6861.
- BARAKAT, A., G. MATASSI and G. BERNARDI, 1998 Distribution of genes in the genome of *Arabidopsis thaliana* and its implications for the genome organization of plants. *Proc. Natl. Acad. Sci. USA* **95**: 10044–10049.
- BARAKAT, A., P. GALLOIS, M. RAYNAL, D. MESTRE-ORTEGA, C. SALLAUD *et al.*, 2000 The distribution of T-DNA in the genomes of transgenic *Arabidopsis* and rice. *FEBS Lett.* **471**: 161–164.
- BECKER, J., and M. HEUN, 1995 Barley microsatellites: allele variation and mapping. *Plant Mol. Biol.* **27**: 835–845.
- CHOI, H. W., P. G. LEMAUX and M. J. CHO, 2002 Use of fluorescence *in situ* hybridization for gross mapping of transgenes and screening for homozygous plants in transgenic barley (*Hordeum vulgare* L.) *Theor. Appl. Genet.* **106**: 92–100.
- CHRISTENSEN, A. H., and P. H. QUAIL, 1996 Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. *Transgenic Res.* **5**: 213–218.
- FORSBACH, A., D. SCHUBERT, B. LECHTENBERG, M. GILS and R. SCHMIDT, 2003 A comprehensive characterization of single-copy T-DNA insertions in the *Arabidopsis thaliana* genome. *Plant Mol. Biol.* **52**: 161–176.
- GILL, K. S., B. S. GILL and T. R. ENDO, 1993 A chromosome region-specific mapping strategy reveals gene-rich telomeric ends in wheat. *Chromosoma* **102**: 374–381.
- GILL, K. S., B. S. GILL, T. R. ENDO and T. TAYLOR, 1996 Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat. *Genetics* **144**: 1883–1891.
- HARWOOD, W. A., S. M. ROSS, P. CILENTO and J. W. SNAPE, 2000 The effect of DNA/gold particle preparation technique, and particle bombardment device, on the transformation of barley (*Hordeum vulgare*). *Euphytica* **111**: 67–76.
- HARWOOD, W. A., S. M. ROSS, S. M. BULLEY, S. TRAVELLA, B. BUSCH *et al.*, 2002 Use of the firefly luciferase gene in a barley (*Hordeum vulgare*) transformation system. *Plant Cell Rep.* **21**: 320–326.
- KLEINHOF, A., A. KILIAN, M. A. SAGHAI MAROOF, R. M. BIYASHEV, P. HAYES *et al.*, 1993 A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome. *Theor. Appl. Genet.* **86**: 705–712.
- LAURIE, D. A., J. W. SNAPE and M. D. GALE, 1992 DNA marker techniques for genetic analysis in barley, pp. 115–132 in *Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology*, edited by P. R. SHEWRY. CAB International, Oxford.
- LAURIE, D. A., N. PRATCHETT, K. M. DEVOS, I. J. LEITCH and M. D. GALE, 1993 The distribution of RFLP markers on chromosome 2(2H) of barley in relation to the physical and genetic location of 5S rDNA. *Theor. Appl. Genet.* **87**: 177–183.
- LAURIE, D. A., N. PRATCHETT, J. H. BEZANT and J. W. SNAPE, 1995 RFLP mapping of 5 major genes and 8 quantitative trait loci controlling flowering time in a winter \times spring barley (*Hordeum vulgare* L.) cross. *Genome* **38**: 575–585.
- LEGGETT, M. J., S. J. PERRET, J. HARPER and P. MORRIS, 2000 Chromosome localization of cotransformed transgenes in the hexaploid cultivated oat *Avena sativa* L. using fluorescence *in situ* hybridization. *Heredity* **84**: 46–53.
- LEITCH, I. J., and J. S. HESLOP-HARRISON, 1993 Physical mapping of four sites of 5S rDNA sequences and one site of the alpha-amylase-2 gene in barley (*Hordeum vulgare*). *Genome* **36**: 517–523.
- LIU, Z.-W., R. M. BIYASHEV and M. A. SAGHAI-MAROOF, 1996 Development of simple sequence repeat DNA markers and their integration into barley linkage map. *Theor. Appl. Genet.* **93**: 869–876.
- MICHELMORE, R. W., I. PARAN and R. V. KESSELI, 1991 Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic

- regions by using segregating populations. *Proc. Natl. Acad. Sci. USA* **88**: 9828–9832.
- PEDERSEN, C., J. ZIMNY, D. BECKER, A. JÄHNE-GÄRTNER and H. LÖRZ, 1997 Localization of introduced genes in the chromosomes of transgenic barley, wheat and triticale by fluorescent *in situ* hybridization. *Theor. Appl. Genet.* **94**: 749–757.
- QIN, H., Y. DONG and A. G. VON ARNIM, 2003a Epigenetic interactions between *Arabidopsis* transgenes: characterization in light of transgene integration sites. *Plant Mol. Biol.* **52**: 217–231.
- QIN, G., D. KANG, Y. DONG, Y. SHEN, L. ZHANG *et al.*, 2003b Obtaining and analysis of flanking sequences from T-DNA transformants of *Arabidopsis*. *Plant Sci.* **165**: 941–949.
- RAMSAY, L., M. MACAULAY, S. DEGLI IVANISSEVICH, K. MACLEAN, L. CARDLE, *et al.*, 2000 A simple sequence repeat-based linkage map of barley. *Genetics* **156**: 1997–2005.
- SALLAUD, C., D. MEYNARD, J. VAN BOXTEL, C. GAY, M. BÈS *et al.*, 2003 Highly efficient production and characterization of T-DNA plants for rice (*Oryza sativa* L.) functional genomics. *Theor. Appl. Genet.* **106**: 1396–1408.
- SALVO-GARRIDO, H. G., S. TRAVELLA, T. SCHWARZACHER, W. A. HARWOOD and J. W. SNAPE, 2001 An efficient method for the physical mapping of transgenes in barley using *in situ* hybridization. *Genome* **44**: 104–110.
- SHA, Y., S. LI, Z. PEI, L. LUO, Y. TIAN *et al.*, 2004 Generation and flanking sequence analysis of a rice T-DNA tagged population. *Theor. Appl. Genet.* **108**: 306–314.
- SPERTINI, D., D. BÉLIVEAU and G. BELLEMARE, 1999 Screening of transgenic plants by amplification of unknown genomic DNA flanking T-DNA. *Biotechniques* **27**: 308–314.
- SVITASHEV, S., E. ANANIEV, W. P. PAWLOWSKI and D. A. SOMERS, 2000 Association of transgene integration sites with chromosome rearrangements in hexaploid oat. *Theor. Appl. Genet.* **100**: 872–880.
- TINGAY, S., D. MCELROY, R. KALLA, S. FIEG, M. WANG *et al.*, 1997 *Agrobacterium tumefaciens*-mediated barley transformation. *Plant J.* **11**: 1369–1376.

Communicating editor: T. BROWN

